

## Informatyczny kącik olimpijski (125):

### Liczby prawie pół-pierwsze

Tym razem omówimy zadanie *Liczby prawie pół-pierwsze*, które pojawiło się na Sobotnim Kole Naukowym (25.11.2017) organizowanym przez Stowarzyszenie Talent w III Liceum Ogólnokształcącym w Gdyni.

**Liczby prawie pół-pierwsze:** Liczbę  $p$  nazywamy prawie pół-pierwszą, jeśli jest iloczynem dwóch liczb pierwszych  $q_1, q_2$ , gdzie  $2 \leq q_1, q_2 \leq 10^6$ . Liczby  $14 = 2 \cdot 7$  i  $49 = 7 \cdot 7$  są prawie pół-pierwsze. Natomiast liczby  $30 = 2 \cdot 3 \cdot 5$  i  $2\,000\,000\,014 = 2 \cdot 1\,000\,000\,007$  nie są prawie pół-pierwsze. Dane są dwie liczby całkowite  $a$  i  $b$  ( $2 \leq a \leq b \leq 10^{12}$ ). Oblicz, ile jest liczb prawie pół-pierwszych na przedziale  $[a, b]$ .

**Rozwiązanie**  $O((b-a) \cdot \sqrt{b})$

Pierwszy pomysł, który nasuwa się po przeczytaniu treści, polega na przejrzaniu wszystkich liczb całkowitych z przedziału  $[a, b]$  i zliczeniu tych, które są prawie pół-pierwsze. Liczba prawie pół-pierwsza zawiera dokładnie dwie liczby pierwsze nie większe niż  $10^6$  w rozkładzie na czynniki pierwsze. Rozkład na czynniki pierwsze liczby całkowitej  $x$  możemy znaleźć w czasie  $O(\sqrt{x})$  (przeładowujemy potencjalne dzielniki pierwsze z przedziału  $[2, \lceil \sqrt{x} \rceil]$ ). Rozwiązanie działa w czasie  $O((b-a) \cdot \sqrt{b})$ .

**Rozwiązanie**  $O(b \cdot \log(\log(b)))$

W tym podejściu, podobnie jak wcześniej, dla każdej liczby całkowitej z przedziału  $[a, b]$  sprawdzimy, czy jest ona prawie pół-pierwsza.

Zacznijmy od stworzenia takiej tablicy  $S$ , że  $S[i]$  dla każdego  $i \in [2, b]$  oznacza najmniejszy dzielnik pierwszy  $i$ . W tym celu wystarczy nieznacznie zmodyfikować sito Eratostenesa. Zamiast zapamiętywać informację, czy liczba jest pierwsza czy złożona, będziemy pamiętali jej najmniejszy dzielnik, będący liczbą pierwszą. Strukturę budujemy w czasie  $O(b \cdot \log(\log(b)))$ . Wówczas w czasie stałym możemy stwierdzić, że:

- $p$  jest pierwsze, jeśli:  $S[p] = p$ ,
  - $p$  jest prawie pół-pierwsze, jeśli:  $S[p] \neq p \wedge S[\frac{p}{S[p]}] = \frac{p}{S[p]} \wedge S[p] \leq 10^6$ .
- Wtedy jej dzielnikami pierwszymi są  $S[p]$  i  $\frac{p}{S[p]}$ .

Po wygenerowaniu  $S$  zliczamy liczby prawie pół-pierwsze z przedziału  $[a, b]$ . Rozwiązanie działa w czasie  $O(b \cdot \log(\log(b)))$ .

**Rozwiązanie**  $O((b-a) \cdot \log(\log(b-a) + \sqrt{b}))$

Niech  $M = 10^6$  oraz  $T[i]$  dla każdego  $i \in [2, M]$  wskazuje, czy  $i$  jest pierwsze ( $T[i] = 1$ ), czy złożone ( $T[i] = 0$ ). Dla pozostałych  $i$  niech  $T[i] = 0$ .  $T$  obliczamy za pomocą sita Eratostenesa. Dodatkowo stwórzmy taką tablicę  $S'$ , że  $S'[i]$  dla każdego  $i \in [a, b]$  będzie oznaczało najmniejszy dzielnik pierwszy  $i$ . Konstrukcja  $S'$  jest następująca. Najpierw przypisujemy  $S'[i] = i$  dla każdego  $i \in [a, b]$ . Następnie dla każdej liczby pierwszej  $q \in [2, \lceil \sqrt{b} \rceil]$  przeładowujemy jej wielokrotności w przedziale  $[a, b]$ , czyli liczby  $\lceil \frac{a}{q} \rceil \cdot q, (\lceil \frac{a}{q} \rceil + 1) \cdot q, \dots, \lfloor \frac{b}{q} \rfloor \cdot q$ , i aktualizujemy ich najmniejszy dzielnik pierwszy. Zauważmy, że wielokrotności 2 w przedziale  $[a, b]$  jest co najwyżej  $\frac{b-a}{2} + 1$ , wielokrotności 3 jest co najwyżej

$\frac{b-a}{3} + 1$  itd. Zatem konstrukcja struktury zajmuje  $O((b-a) \cdot \log(\log(b-a) + \sqrt{b}))$  operacji.

Na koniec zliczamy liczby prawie pół-pierwsze z przedziału  $[a, b]$ , czyli takie  $p \in [a, b]$ , że:

$$S'[p] \neq p \wedge S'[p] \leq M \wedge T\left[\frac{p}{S'[p]}\right] = 1.$$

**Rozwiązanie**  $O(M \cdot \log(\log(M)))$

Niech  $f: \mathbb{N} \rightarrow \mathbb{N}$  będzie taką funkcją, że  $f(x)$  oznacza liczbę liczb prawie pół-pierwszych na przedziale  $[1, x]$ . Wówczas liczb prawie pół-pierwszych na przedziale  $[a, b]$  jest  $f(b) - f(a-1)$ .

Opiszemy teraz, jak policzyć  $f(x)$  dla  $x \in [1, 10^{12}]$ . Zacznijmy od wygenerowania ciągu kolejnych liczb pierwszych nie większych niż  $M = 10^6$ . Oznaczmy ten ciąg jako  $P = (p_1, p_2, \dots, p_k)$ . Początkowymi wyrazami ciągu są:  $p_1 = 2, p_2 = 3, p_3 = 5$  itd. Liczby pierwsze możemy znaleźć za pomocą sita Eratostenesa. Zauważmy, że tylko elementy ciągu  $P$  mogą występować w rozkładzie na czynniki pierwsze liczb prawie pół-pierwszych. Zatem:

$$f(x) = |\{(i, j) : p_i p_j \leq x \wedge 1 \leq i \leq j \leq k\}|.$$

Wartość  $f(x)$  możemy policzyć naiwnie, przeglądając wszystkie pary elementów ciągu  $P$ . Niestety takie rozwiązanie jest wolne, ponieważ przegląda się  $O(k^2)$  par elementów, gdzie  $k = 78\,498$  (tyle jest liczb pierwszych nie większych niż  $10^6$ ).

Na szczęście możemy to przyspieszyć. Niektóre pary będziemy zliczali „hurtowo”. Dokładniej, dla każdego  $i \in [1, k]$  wyznaczmy takie największe  $j_i \in [i, k]$ , że  $p_i p_{j_i} \leq x$ . Jeśli takie  $j_i$  nie istnieje, to niech  $j_i = i - 1$ . Wówczas liczby postaci  $p_i p_i, p_i p_{i+1}, \dots, p_i p_{j_i}$  są prawie pół-pierwsze. Zaś wszystkich par jest  $\sum_{i \in [1, k]} (j_i - i + 1)$ . Algorytm przebiega następująco. Najpierw naiwnie znajdujemy  $j_1$ , przeglądając przedział indeksów  $[1, k]$ . Następnie kolejno wyznaczamy wartości  $j_2, j_3, \dots, j_k$ .

**Obserwacja:** Dla każdego  $i \in [2, k]$  zachodzi  $j_i \leq j_{i-1}$ .

**Dowód:** Wiemy, że  $p_i > p_{i-1}$  oraz  $\forall_{l > j_{i-1}} p_{i-1} p_l > x$ . Stąd  $\forall_{l > j_{i-1}} p_i p_l > x$ . Zatem  $j_i \leq j_{i-1}$ .

Na mocy powyższej obserwacji, szukanie wartości  $j_i$  możemy rozpocząć od  $j_{i-1}$  i zmniejszać ją dopóki  $p_i p_{j_i} > x$ . W ten sposób wykonamy  $O(k)$  operacji podczas obliczania  $f(x)$ . Całe rozwiązanie działa w czasie  $O(M \cdot \log(\log(M)))$ .

Bartosz ŁUKASIEWICZ